(12) UK Patent Application (19) GB (11) 2 353 926 (13) A

(43) Date of A Publication 07.03.2001

(21) Application No 9920811.8

(22) Date of Filing 04.09.1999

(71) Applicant(s)
Central Research Laboratories Limited
(Incorporated in the United Kingdom)
Dawley Road, HAYES, Middlesex, UB3 1HH, United Kingdom

(72) Inventor(s)
Michael Percy
Alastair Sibbald

(74) Agent and/or Address for Service
QED I.P. Services Ltd
Dawley Road, HAYES, Middlesex, UB3 1HH, United Kingdom

(51) INT CL^7
H04S 5/00

(52) UK CL (Edition S)
H4R RSS R16A2

(56) Documents Cited
WO 98/30064 A   US 5555306 A   US 4625326 A

(58) Field of Search
UK CL (Edition R) H4R RSS RSX
INT CL^7 H04R 3/00 5/00 5/04 , H04S 1/00 3/00 5/00 5/02
online: EPODOC,WPI,JAPIO

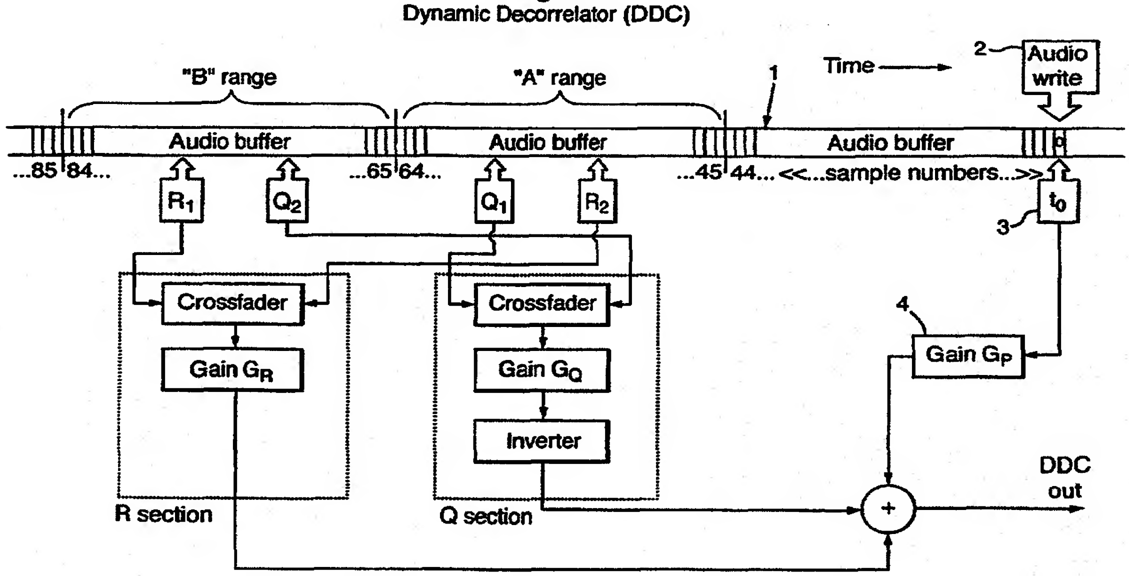
(54) Abstract Title

Generating a second audio signal from a first audio signal for the reproduction of 3D sound

(57) A method of generating a second decorrelated audio signal from a first audio signal 2, for use in synthesising a 3D sound field, includes :- a) deriving from the first signal a first delayed signal using an audio delay line 1; b) multiplying this first delayed signal by a gain factor G_Q between zero and minus one to give a first delayed gain-adjusted signal; c) deriving from the first audio signal a second delayed signal, having a different delay time from the first delayed signal; d) multiplying this second delayed signal by a gain factor G_R between zero and plus one (such that the said gain factors sum to zero) to give a second delayed gain-adjusted signal; e) combining said first and said second delayed gain-adjusted signals with the first audio signal to provide a second decorrelated audio signal DDC. The first and second delayed signals are delayed by time periods which change in a substantially random manner. Each of the first and second delayed signals may be derived from two taps Q_1,Q_2 and R_1,R_2 of the audio delay line 1, the signals from the taps being crossfaded.

Fig.3.

Dynamic Decorrelator (DDC)



6

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

GB 2 353 926 A

Fig.1.

Schematic representation of simple comb-filter (upper), and resultant characteristics for time delay = 1 ms and attenuation factor = 0.5

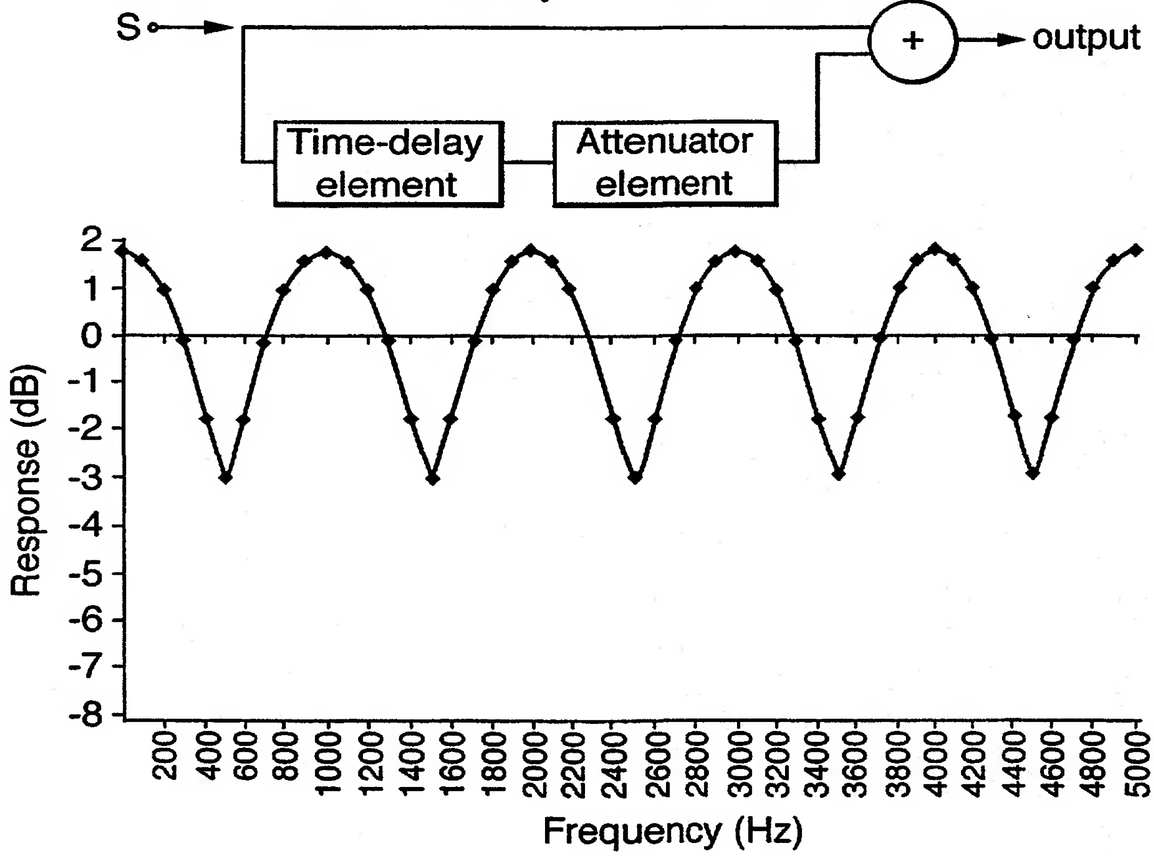


Fig.2.

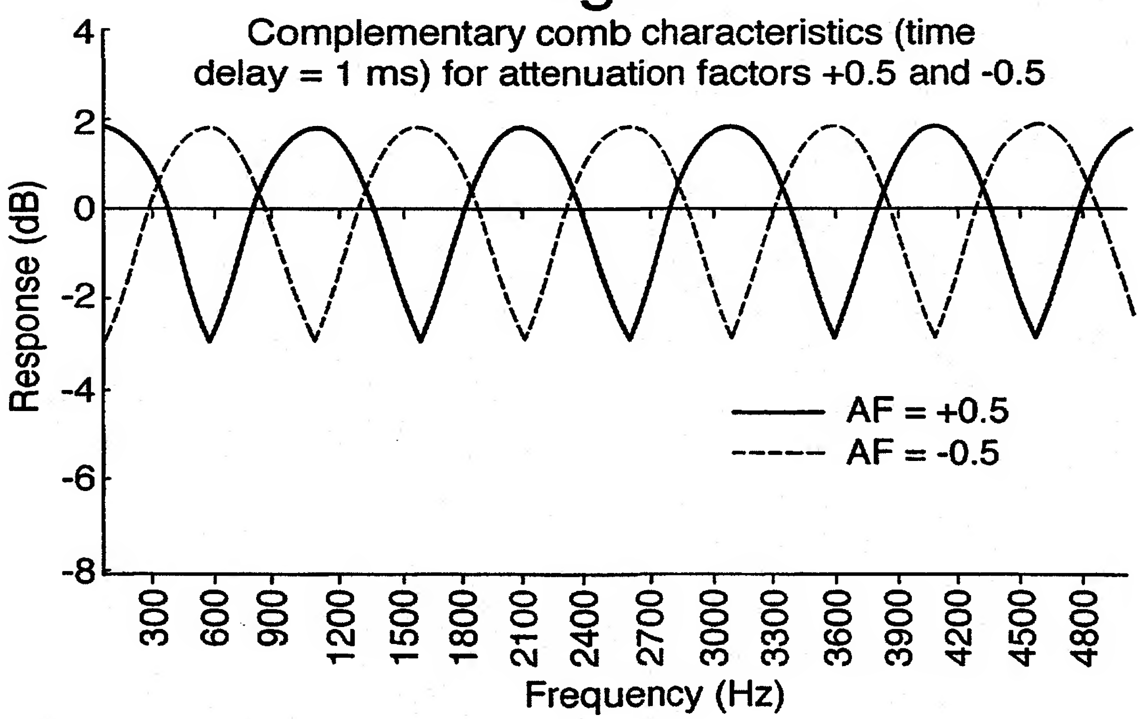


Fig.3.
Dynamic Decorrelator (DDC)

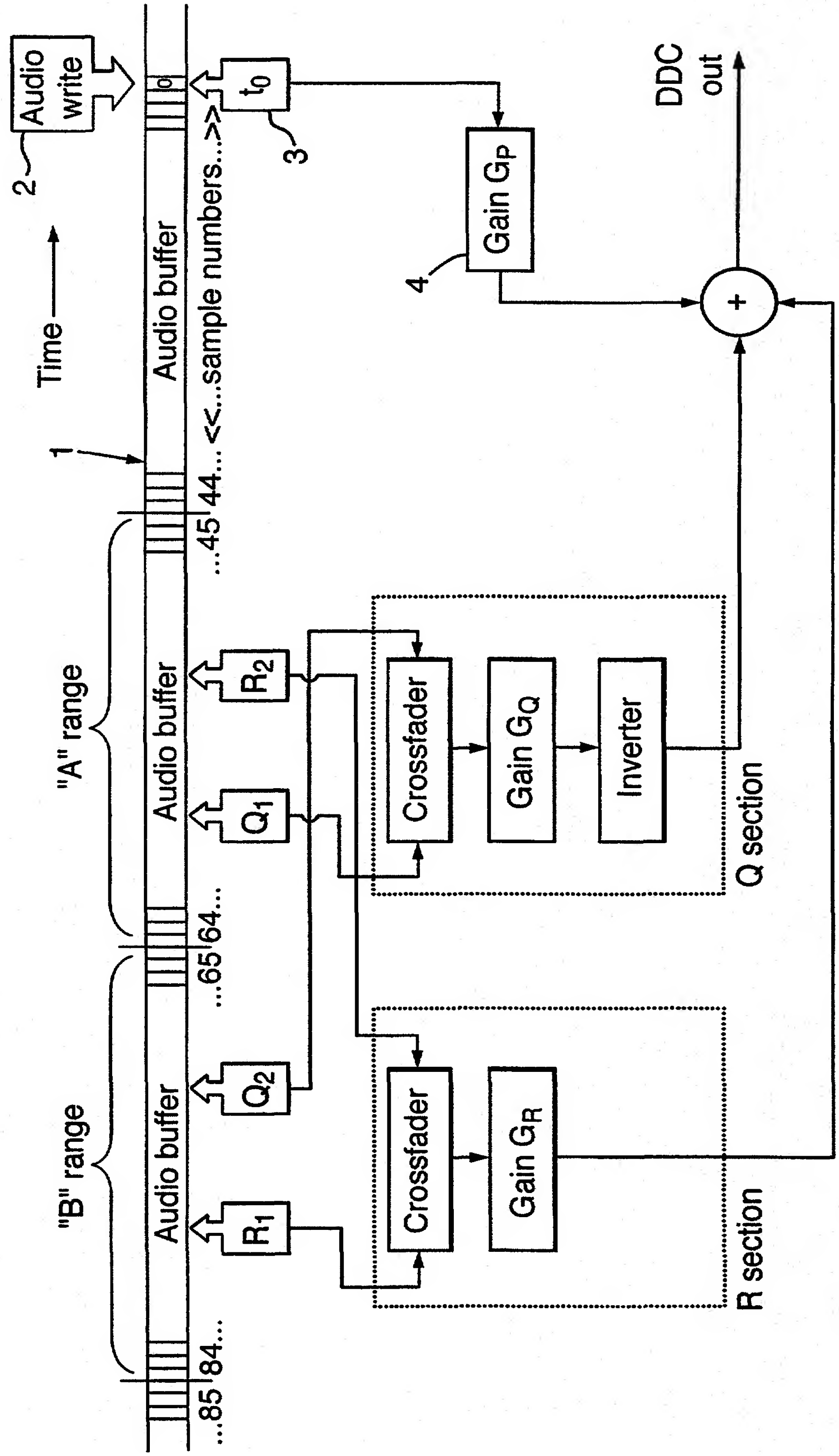


Fig.4.

Schematic diagram of system dynamics

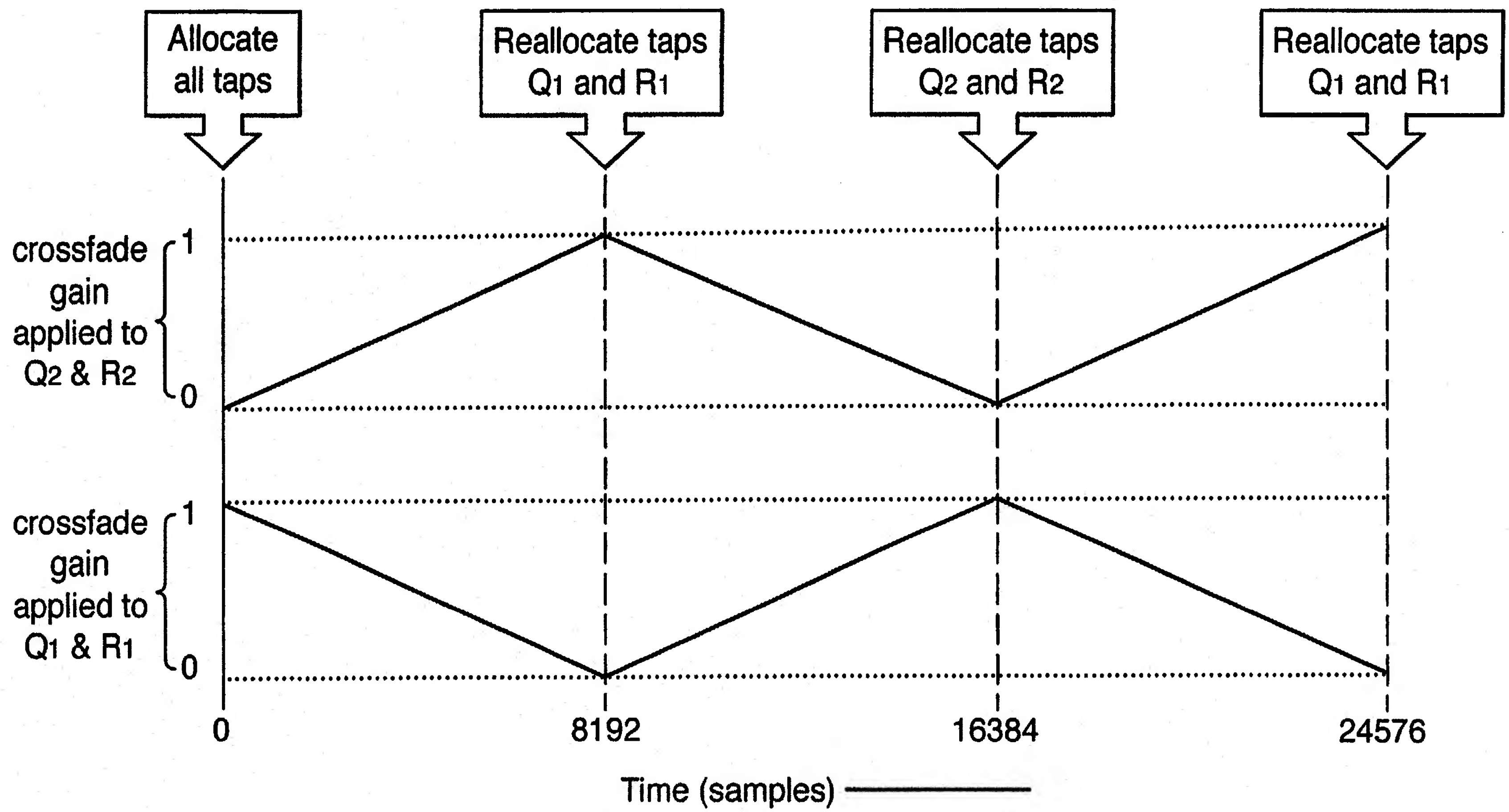


Fig.5.

Decorrelated amplitude spectrum

Amplitude spectrum for $R_1 = 78$ and $Q_1 = 47$

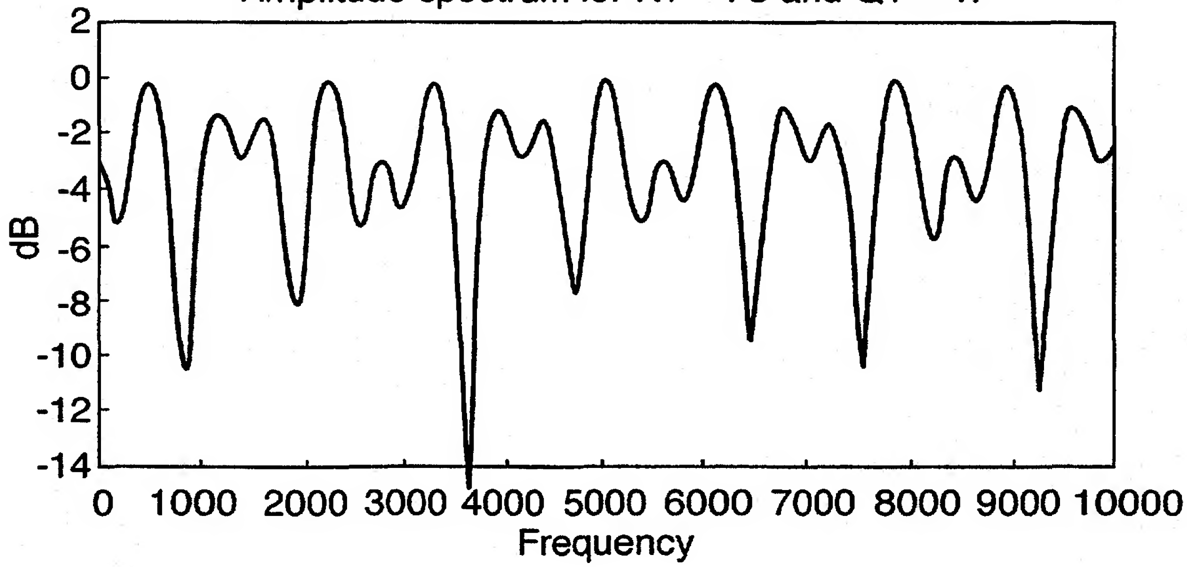
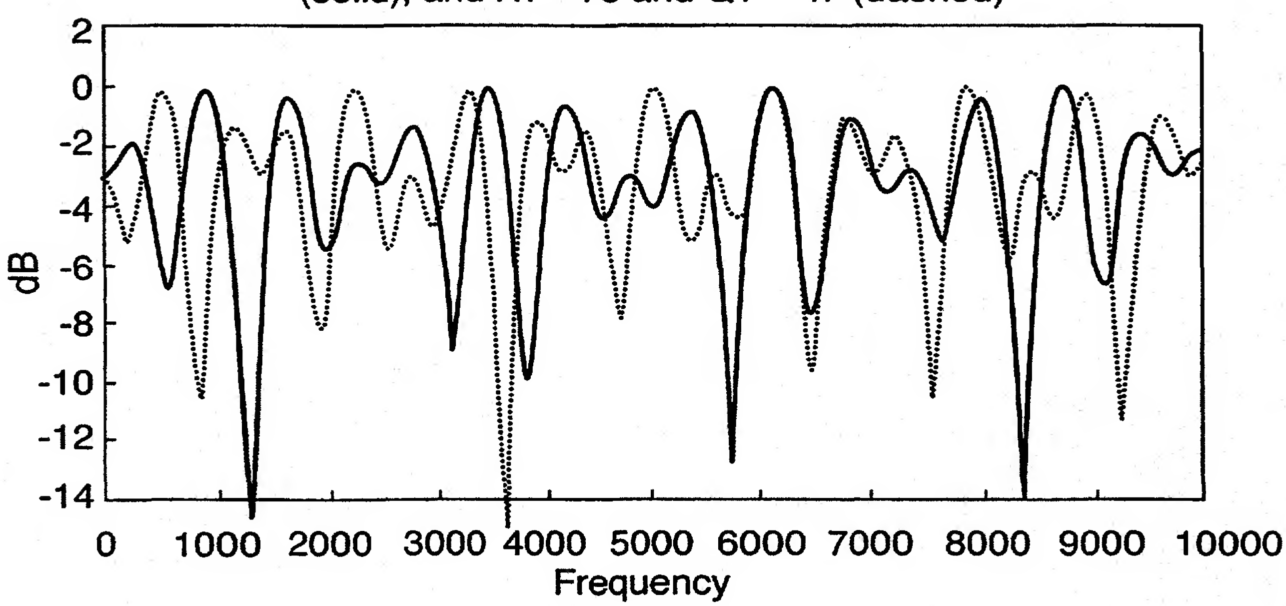


Fig.6.

Decorrelated amplitude spectra for 2 pairs of Q and R tap values

Amplitude spectra for $R_2 = 50$ and $Q_2 = 68$ (solid), and $R_1 = 78$ and $Q_1 = 47$ (dashed)



METHOD AND APPARATUS FOR GENERATING A SECOND AUDIO SIGNAL FROM A FIRST AUDIO SIGNAL

This invention relates to the reproduction of 3D-sound from two-speaker stereo systems, headphones, and multi-speaker audio systems. It relates particularly, though not exclusively, to a method for the creation of one or more virtual sound sources simultaneously from a single, common sound signal which, nevertheless, can be discerned separately from one another by a listener in use.

Such methods have been described in general terms in US 5,666,425, WO98/52382, and our co-pending UK patent applications GB9813290.5 and GB9905872.9, which are incorporated herein by reference. The latter contains a comprehensive description of how head-related transfer functions (HRTFs) are used in the synthesis of 3D sound.

Technical Background

The Haas (or Precedence) Effect [M B Gardner, J. Acoust. Soc. Am., 43, (6), pp.1243-1248 (1968)] is the phenomenon that the brain, when presented with several similar pieces of audio information at slightly differing times to process, uses only the first information to arrive from which to compute directional information. The brain then attributes the subsequent, similar information packets with the same directional information. The key to this is that the brain recognises signals which are related to one another (i.e. correlated), and processes them in a particular way.

For example, if several loudspeakers play music in a room, each at exactly the same loudness, it would appear that all the sound comes from the nearest loudspeaker, and that all the others would appear to be silent. The first sounds to arrive at the listener are used to determine the spatial position of the sound source, and the subsequent sounds simply make it appear louder. This effect is so strong that the intensity of the second signal could be up to 8 dB greater than the

initial signal, and the brain would still use the first (but quieter) signal to decide where the sound originated.

This effect is also known under the names "law of the first wavefront", "auditory suppression effect", "first-arrival effect" and "threshold of extinction", and it is used for the basis of sound reinforcement used in Public Address systems.

The brain attributes great relative importance to time information as opposed to intensity information. For example, an early paper by Snow [W B Snow, J. Acoust. Soc. Am., 26, (6), pp.1071-1074 (1954)] describes experiments on compensating differences in left-right intensity balance using relative L-R time delays. It was reported that a 1 ms time delay would balance as much as 6 dB of intensity misbalance.

It seems possible that this mechanism has evolved so that the brain can deal with multiple reflections in a room without confusion. This would enable the rapid location of the primary sound-source, distinguished clearly from a confusing array of first-order sound images caused by the reflections.

The relevance of the precedence effect here is that it can contribute to "spatial fusion", under particular circumstances, during the synthesis of virtual 3D-sound images. The "Spatial Fusion" effect is not widely appreciated or known, and it is common both to loudspeaker and headphone listening. It occurs when synthesising several virtual sound sources from a single, common source, such that there is a significant common-mode content present in the left and right channels (and rear-left and rear-right channels in a four-speaker 3D-audio system).

Example 1. Primary signal + derived reverberation.

When 3D "sound-scapes" are created from many individual sound-sources, any signals which have been derived directly from another sound source (such as a [secondary] reverberation signal created from a primary source), are perceived to "combine" spatially with the primary signal if they are presented to the listener within a period of about 15 ms. Beyond this time period, they begin to be

discernible as separate entities, in the form of an echo. The effect of such spatial combination is to inhibit the secondary image and create an imprecise and vaguely positioned spatial image at the location of the primary sound source.

Example 2. Symmetrical placement with a common-mode signal present.
In some circumstances, especially when virtual sound effects are to be recreated to the sides of the listener, the HRTF processing decorrelates the individual signals sufficiently such that the listener is able to distinguish between them, and hear them as individual sources, rather than "fuse" them spatially into apparently a single sound. However, when a pair (or other, even number) of virtual sounds are to be synthesised in a symmetrical configuration about the median plane (the vertical plane which bisects the head of the listener, running from front to back), the symmetry enhances any correlation between the individual sound sources, and the result is that the perceived sounds can become spatially "fused" together into one. For example, if it is required to "virtualise stereo" for headphone listening (i.e. create virtual left- and right-sources at azimuth angles ±30° for the respective channels), then this can be achieved reasonably well using discrete, individual virtual sources. However, if a stereo music source were used, then, inevitably, the centrally positioned elements of the stereo mix would present a significant common-mode signal, and so the perceived virtual sound image would tend to collapse. Instead of appearing as a pair of external sound sources, the sound image would become centralised and perceived to be "inside the head" for headphone users.

These limitations are caused and exacerbated by the unnaturally high degree of correlation between the signals which are presently used in 3D-sound synthesis. The situation is also much less true-to-life, in that the virtual sound emitters are implemented as simplistic "point" sources, rather than as line or area sound-emitters. (Methods for remedying this have been described in our co-pending patent application GB9813290.5.)

In reality, a line or area sound-emitter can be considered to be the sum of many individual elemental sound-sources which all possess differing amplitude and phase characteristics. In a static, real-world environment, there are usually many objects and surfaces asymmetrically placed about the listener, locally, which scatter and reflect the sound waves differently on their paths to the left and right ears of the listener. In other words, there is a degree of decorrelation occurring between the originally emitted sound and the sum of the elemental components when they arrive at the listener's ears. In a "dynamic" environment, in which there is also relative movement between the emitter(s) and listener, the integral sum of the phase and amplitude characteristics perceived by the listener are constantly changing, and hence the perceived signals are, again, decorrelated with respect to the originally emitted signal, and the decorrelation properties are changing dynamically. This is further enhanced by the changing contributions from the locally scattered and reflected waves. These effects reduce the amount of amplitude and phase correlation between:

a. the signal from one single sound-source, as measured at different points in space (e.g. the left and right ears); and also

b. identical signals emitted from two (or more) symmetrically placed sources, such as a loudspeaker pair, measured at one central point in space (and, of course, at different points in space).

(It is worth noting that the tonal properties of the perceived sounds are relatively unaffected by these processes.)

In summary, perceived sounds in the real world undergo natural decorrelation with respect to the original source. In a moving environment, the decorrelation parameters are changing dynamically.

In practise, however, usually for the sake of economy of storage and processing, we are often obliged to use only a single sound recording or computer file from which to create a plurality of virtual sources. Consequently, there is an unrealistically high correlation between the resultant array of virtual sources: a unique and unnatural situation. This makes the sound image susceptible to spatial fusion and collapse. It is the recognition of this process, and the

description of a method of synthesising apparently naturally decorrelated sound signals which forms the basis of the present invention.

One important system which has become an industry "standard" for many consumer products related to home cinema and digital TV, is the Dolby Pro-Logic (RTM) system, or Dolby Surround (RTM). It is characterised by the encoding of four channels of analogue audio into two channels of analogue audio, such that it can be recorded on to video tapes and DVDs (and also used for TV broadcast), from which the signals are decoded and used to drive four loudspeakers (left, centre, left-surround and right-surround), together with an optional sub-woofer. However, the bandwidth limitations only allow a single rear-channel "surround" signal. If this signal was fed in parallel to both rear loudspeakers, the Precedence Effect would make the surround channel audio all appear to come from the nearest loudspeaker only. In order to make the surround channel seem more spacious, the surround signal is fed directly to one of the rear speakers, but it is inverted before being sent to the other rear loudspeaker. This is a crude way to decorrelate the signals being emitted from both surround speakers, but it assists the listeners to perceive sounds emanating from both loudspeakers, rather then just one, thus creating a more spacious experience. Of course, there can be no rear sound image formed by this means, only spatial effects to enhance the frontal sound images and create "surround" sound.

Two important new applications for the virtualisation of Dolby Surround (RTM) material are (a) the playback of DVD movies on multimedia systems (via loudspeakers); and (b) the provision of headphone virtualisation for home cinema systems for "quiet" late-night listening. The problems here are: (a) how might it be possible to generate two separately perceivable surround channels from a single source; (b) how might it be possible to prevent the centre channel (which is entirely common to left and right channels) from collapsing the sound image; and (c) how can reverberation be generated and virtualised without fusing the image?

One of the most important applications for 3D audio at present is "3D Positional Audio" processing for computer games. In order to synthesise audio

material bearing 3D-sound cues for the listener to perceive, the signals must be convolved with one or more appropriate HRTFs, and then delivered to the listener's ears in such a way that his or her own hearing processes do not interfere with the in-built 3D cues. This can be achieved either by listening through headphones, or via loudspeakers in conjunction with a suitable transaural crosstalk-cancellation scheme (as described in co-pending patent application GB9816059.1). In order to provide a more realistic experience for the listener, we recently devised a method for creating line and area virtual sound-sources (GB9813290.5). An important feature of that invention is the need to provide one or more signals which have been decorrelated from the primary source. This was achieved by use of one or more comb filters, but it will become evident that there are considerable limitations on the use of comb filters. The present invention, however, is ideally suited for use in this particular application (marketed under the trademark "Sensaura ZoomFX"), in addition to other fields of application.

Prior Art

Pseudo-stereo

A method of creating "pseudo-stereo" has been described in US 4,625,326 in which tapped delay-lines with feed-0back loops were used to create a comb-filtered signal pair. It seems likely that this was intended for portable stereo applications.

Dolby Surround (RTM) Virtualisation

US 5,844,993 discloses use of a complementary comb filter pair to create a pair of rear channels from the single "surround" channel, and shows the first notch and peak features occurring at 100 Hz.

Sensaura (RTM) ZoomFX

The use of comb-filtering was described in GB9813290.5, but it is worth re-stating below in order to establish the basic method and typical results.

Comb Filters

A signal can be decorrelated by means of comb-filtering, as is known in the prior art. Figure 1 shows a simple comb filter, in which the source signal, S, is passed through a time-delay element, and an attenuator element, and then combined with the original signal, S. At frequencies where the time-delay corresponds to one half a wavelength, then the two combining waves are exactly 180° out of phase, and cancel each other, whereas when the time delay corresponds to one whole wavelength, the waves combine constructively. If the amplitudes of the two waves are the same, then total nulling and doubling, respectively, of the resultant wave occurs. By attenuating one of the combining signals, as shown, then the magnitude of the effect can be controlled. For example, if the time delay is chosen to be 1 ms, then the first cancellation point exists at 500 Hz. The first constructive addition frequency points are at 0 Hz, and 1 kHz, where the signals are in phase. If the attenuation factor is set to 0.5, then the destructive and constructive interference effects are restricted to –3 dB and +3 dB respectively. These characteristics are shown in Figure 1 (lower).

It might be often required to create a pair of decorrelated signals. For example, when a large sound source is to be simulated in front of the listener, extending laterally to the left and right, a pair of sources would be required for symmetrical placement (e.g. -40° and +40°), but with both sources individually distinguishable. This can be done by creating a pair of complementary comb filters. This is achieved, firstly, by creating an identical pair of filters, each as shown according to Figure 1 (and with identical time delay values), but with signal inversion in one of the attenuation pathways. Inversion can be achieved either by (a) changing the summing node to a "differencing" node (for signal subtraction), or (b) inverting the attenuation coefficient (e.g. from +0.5 to -0.5); the end result is the same in both cases. The output of such a pair of complementary filters exhibits maximal amplitude decorrelation within the constraints of the attenuation factors, because the peaks of one correspond to the troughs of the other (Figure 2), and vice versa.

If a source "triplet" were required, then a convenient method for creating such an arrangement is the creation of a pair of maximally decorrelated sources, which are then used in conjunction with the original source itself, thus providing three sources.

Problems with Comb Filters

There are three significant problems associated with the use of comb filters to process audio, as follows.

Audible artefacts. As can be seen in Figure 1, the property of a comb filter is to create a series of notches and peaks throughout the spectrum, with the frequency of the lowest feature determined by the time-delay of the filter. Our hearing processes are particularly good at noticing notches in the audio spectrum, and we are also good at detecting tones and notches which are repeated at octave intervals (where the frequencies are multiple values of a fundamental value). Consequently, a comb-filtered signal sounds very artificial, tonally.

Doppler interaction. When more than one comb-filtered signals are subjected to Doppler-effect type processing (as happens in computer game audio applications), then the comb artefacts in the audio become exaggerated, apparently by the interaction between the comb features. Even if one uses complementary comb-filters to make the sources, as described above, the Doppler processing can shift the features in the frequency domain such that they "slide" over each other and become noticeable as artefacts. Notches which are caused to "move" in the frequency domain are especially noticeable: a good example of this is the "flangeing" effect used for music-effects processors, and another is the effect which is heard as a steam train arrives, hissing, at the station platform. The hiss sound is, approximately, a form of white noise, and it arrives at the listener both directly and also reflected from the platform surface where it combines with the direct path sound. The time-delay difference between the two is small when the train is distant, but increases to correspond to a path length of about twice the listener's ear height above ground when the sound source is above the listener's head and close. For example, if the train were about 4 m distant (with an elevated

source), and the ear height were 1.8 m, then the delay would be about 4 ms, and so the first (lowest) notch would occur at about 125 Hz.

Limits on the number of processed channels. If more than two or three decorrelated sources from a single monaural source are required, then there are problems in creating a sufficiently large number of filtering options because their properties would overlap significantly. Consequently, the amount of decorrelation would diminish and the effectiveness would be reduced.

The Invention

According to a first aspect of the invention, there is provided a method as claimed in claims 1 - 3. According to a second aspect of the invention, there is provided an apparatus as specified in claims 4 - 8.

The present invention is a means of decorrelating a sound source so as to provide one or more sound sources which can be used for 3D-sound synthesis such that they can be perceived independently of one another. The invention is advantageous over the use of simple comb-filtering, in that: (a) there are no significant audible artefacts present; (b) the derived sources can be Doppler processed without flangeing artefacts, and (c) a plurality of sources can be derived from one single source.

Embodiments of the invention will now be described, by way of example only, with reference to the accompanying schematic drawings, in which:-

Figure 1 shows a prior art comb filter and characteristic,

Figure 2 shows the outputs of a pair of complementary comb filters

Figure 3 shows a schematic representation of an apparatus according to the invention,

Figure 4 shows a schematic representation of the dynamic operation of the apparatus of Figure 3 with time,

Figure 5 shows an amplitude spectrum of a decorrelated resultant signal at one point in time, and

Figure 6 shows a pair of decorrelated resultant signals having different output tap values corresponding to a different point ion time, superimposed on the spectrum of Figure 5.

An embodiment of the present invention in the form of a dynamic decorrelator is shown schematically in Figure 3. It can, of course, be implemented in software or hardware forms.

It includes an audio delay-line (1), which is tapped at two (or more) points within a prescribed range, said points changing frequently and randomly. The outputs of the tap nodes are multiplied by predetermined gain factors, one of which is negative, and then added to the original signal. The effect of this is to cause the spectral profile of the derived signal to change, continually, with respect to the original (and, similarly, there are continual changes in relative phase).

Audio buffer

The central feature is an audio delay line (1) in the form of a buffer, as shown at the top of Figure 3, to which audio is written via the "audio write" pointer (2). The current data byte is read via the "t_0" pointer (3). The "audio write" pointer (and all the data pointers) moves incrementally towards the right after each sample has been written. (An alternative way to view this process is that, effectively, the audio data is injected into the buffer via "audio write", and all the audio data is incrementally streamed leftwards by one cell per sample, flowing past the pointers.) Typically, the audio sampling rate will be 44.1 kHz, and hence the corresponding sampling period is about 22.68 μs. There are two time-delay ranges defined in the audio buffer: an "A" range, encompassing sample numbers 45 to 64 inclusively, and a "B" range, encompassing sample numbers 65 to 84 inclusively.

Read/write pointers

As has been described, there is an "audio write" pointer, via which the audio data is written to the buffer, and a "t_0" pointer, via which the present data byte is read.

There are also four additional "read" pointers, designated R_1, R_2, Q_1 and Q_2. These feed data to the Q and R processing blocks (below). The R_1 and Q_2 pointers always lie in the "B" range of the buffer, and the R_2 and Q_1 pointers always lie in the "A" range of the buffer. The allocation of their positions is changed frequently, as will be described.

Processing blocks "Q" and "R"

The processing blocks Q and R both comprise a crossfader and a fixed gain amplifier (or attenuator), and the Q block also contains an inverter. Each crossfader has two audio inputs and a single audio output. One input to each crossfader is connected so as to receive audio data from a read pointer in the "A" range, and the other input is connected so as to receive audio data from a read pointer in the "B" range. Initially, the crossfader is set to transfer signal to the output from one of the inputs with a gain of unity, and from the other input with a gain of zero. These gain factors are controlled so as to progressively reduce the unity-gain factor to zero, and increase the zero-gain factor to unity; this is done incrementally and synchronously with the audio sampling. The effect is to gradually and continuously crossfade the input to the gain stage of the processing block between the two associated "read" pointers. When the crossfade has been completed, the taps which are now the "zero-gain" ones are reallocated within their range, and the next crossfading cycle begins with the crossfading process reversed. The crossfading cycles continue in this way, such that the inputs to the Q and R gain sections are, in effect, continually changing within the "A" and "B" ranges. This is done sufficiently rapidly so as to render the resultant decorrelating amplitude features inaudible, but slowly enough to avoid modulation noise and for the features to work successfully. Typically, a crossfade cycle rate of greater than 0.5 Hz, preferably 5 - 100 cycles per second is chosen, although much higher cycle rates can, in principle, be used. The Q and R processing block gain stages have fixed gain (or attenuation). It is convenient, but not essential, that they are set equal to one another, because the decorrelation contributions from both cells would be equally weighted. It is also convenient that the sum of all the gain

stages (G_P, G_Q and G_R) is unity, because this corresponds to a maximum overall gain of unity (0 dB) through the system with respect to the original audio signal written to the buffer.

<u>Summing and output node</u>

The output from the Q and R sections are fed into a summing node, together with the output from the t_0 "read" pointer, which is transferred to the node via a fixed gain stage, G_P (4). The output of the summing node is the final system output: the dynamically decorrelated signal.

<u>Dynamics of operation</u>

A description of the dynamic operation of the system follows, with reference to Figures 3 and 4.

The system is initialised prior to use: (a) the "write" and "t_0" pointers are allocated; (b) the R_1, R_2, Q_1 and Q_2 pointers are allocated to random locations within their respective ranges (the R_1 and Q_2 pointers always lie in the "B" range of the buffer, and the R_2 and Q_1 pointers always lie in the "A" range of the buffer); (c) the gain of the Q and R gain stages is set; and (d) the Q and R crossfaders are configured such that, initially, the Q_1 and R_1 pointers transfer data to the Q and R gain blocks with unity-gain, and from the Q_2 and R_2 pointers with zero-gain.

The first audio sample is written into the buffer. Data is read from all "read" taps, processed by the associated crossfaders and gain stages, and then summed by the summing (output) node.

The pointers are all shifted by one sample (to the right in Figure 3), ready for the next read/write event, and the crossfaders are incremented. The gain of the zero-gain crossfade path (i.e. Q_2 and R_2 at this point), is increased by a factor of: {crossfade cycle period} x {sample frequency}. For example, if the crossfade cycle period is to be 0.2 s and the sampling frequency were 44.1 kHz, then the crossfade cycle period would be 0.2 x 44,100 = 8,820 samples. In practise,

however, it would be more convenient to use a processing block length of 8192 samples (corresponding to a crossfade cycle rate of about 5.4 per second).

Accordingly, on this basis, the gain factor for the zero-gain crossfade path would be increased from 0 to 1/8192. Similarly, the gain factor for the unit gain (at this point) crossfade path would be decreased from 1 to 8191/8192.

Items 2 and 3 are repeated until the crossfade from Q_1 to Q_2 (and R_1 to R_2) has fully occurred, after 8192 samples (see Figure 4). At this point, the gain contribution from pointers Q_1 and R_1 is zero, and they are reallocated to new, random positions within their specified ranges. The crossfade process is now reversed so as to fade progressively, sample by sample, to these newly allocated taps, such that after another 8192 samples (16384 in all), the unity-gain path is, once again, from pointers Q_1 and R_1, and the zero-gain path from Q_2 and R_2, at which point they are reallocated to new, random positions within their specified ranges. This cyclic process is repeated continually.

<u>Decorrelation effects</u>

The decorrelation effects of the system are best illustrated by considering what occurs at a point in time when the audio buffer is sufficiently full (i.e. more than 85 samples have been written to it) and the crossfade cycle has reached a reversal point. This occurs in Figure 4, for example, after 16384 samples. Let us also assign some locations, randomly, to the Q and R pointers, as follows:

Q_1 : Range "A", positioned @ 47 samples;
Q_2 : Range "B", positioned @ 68 samples;
R_1 : Range "B", positioned @ 78 samples;
R_2 : Range "A", positioned @ 50 samples.

At this point, "read" pointers Q_2 and R_2 have zero-gain contributions, and Q_1 and R_1 have unity-gain contributions. We choose the processing block gain factors to sum to 1 (above), and with G_R and G_Q equal: say

G_P : 0.50;

G_Q : 0.25;

G_R : 0.25.

Under these (minimal) conditions, there are three contributions to the summing node (although, note that virtually all of the time (i.e. during the crossfading), there will be five contributions):

Q_1 positioned @ 47 samples, via G_Q (0.25) and inverter;
R_1 positioned @ 78 samples, via G_R (0.25);
t_0 positioned @ 0 samples, via G_P (0.50).

Consequently, the output signal at this point in time is the sum of three vectors (in contrast to the comb filter described earlier, which is the sum of only two vectors), although it is the sum of five vectors almost all of the time. This introduces a pseudo-random modification of the amplitude and phase spectra, constrained by the chosen parameters. An amplitude spectrum of the resultant signal created by the parameters used in the above example is shown in Figure 5.

Note that the maximum gain is unity (0 dB), which occurs when all three contributions are effectively in phase (taking account of the inverter) and because the gains are 0.50, 0.25 and 0.25. Also note that the spectral profile is somewhat pseudo-randomly aperiodic (albeit not perfectly so), unlike that of a comb filter, which is perfectly regular and periodic. This feature is important because the profiling is much less audible as an artefact, making the overall effect "tone-neutral".

Another important feature is that the low-frequency gain is always the same (-3 dB) for the system whatever tap allocations are assigned to Q_1, Q_2, R_1 and R_2, because the three contributions become:

$$\{+0.50\,(G_P)\} + \{+0.25\,(G_Q)\} + \{-0.25\,(G_R)\} = 0.50$$

This very important for three reasons. Firstly, there is no low-frequency (LF) degradation of the audio, which is important for interactive sound-effects and music; secondly, the system parameters can be changed dynamically without audible artefacts; and thirdly, because this enables a number of these dynamic decorrelators to be operated simultaneously without cross-interference and with equal weighting. For example, if one inspects Figure 2 of US Patent 5,844,993, one can see that the LF gain of one of the surround channels (from the complementary comb filters) tends to zero (and the other to unity), which creates a massive (total) imbalance between the left- and right-surround channels. This is especially detrimental for home-cinema surround-sound applications in which the audio is especially rich in frequencies between 40 Hz and 500 Hz.

Now consider the next stage. As the processing continues, the crossfade cycle gradually transfers the source of the Q and R processing blocks from Q_1 and R_1 to Q_2 and R_2. At 24576 samples the crossfade has been completed (Figure 4), and the contributions are now as follows.

Q_2 positioned @ 68 samples, via G_Q (0.25) and inverter;
R_2 positioned @ 50 samples, via G_R (0.25);
t_0 positioned @ 0 samples, via G_P (0.50).

Once again, of course, the output signal is the sum of the three vectors, but now the pseudo-random modifications of the amplitude and phase spectra are different because of the changed tap locations, as is indicated by the amplitude spectrum shown in Figure 6, which also includes the previous data of Figure 5 to show the differences. (The phase spectra have not been shown here because they are relatively meaningless owing to the "wrap-around" effect which happens when phase differences exceed 2π.)

In summary, the decorrelated spectral profile of Figure 5 has been gradually transformed into the spectral profile of Figure 6 (solid line) in about one

fifth of a second, and it continues to change, smoothly, continuously and randomly, within the constraints of the specified parameters.

The main advantages of this novel method of decorrelation are as follows.

1. The "complementary" method of using five vectors is "tone neutral".
2. The spectral features are continually changing, and therefore not significantly audible.
3. A plurality of decorrelators can be created and operated from the same source, without cross-conflicts, by seeding differently the initial Q and R "read" tap values.
4. There is no LF degradation of the audio.
5. Identical LF convergence ensures smooth transition between crossfade cycles.
6. Identical LF convergence ensures smooth fading transitions between different decorrelators, as would occur in ZoomFX applications.
7. Identical LF convergence ensures perfect LF balance between different decorrelators running from the same source, as would occur for Dolby Surround(RTM) (Pro-Logic) applications.

For the purpose of clarity, only the simplest implementations of the invention have been described here. Clearly the concept could be implemented in more complicated ways (for example, it would be possible to use a greater number of taps in the audio buffers, and additional processing sections (like Q and R).

The specified ranges and rates here are the ones which we are now using, and have been cited purely for example: naturally they could be extended and changed.

The main applications are related to (a) Sensaura (RTM) ZoomFX; (b) the virtualisation of Dolby Digital (RTM) (creating several right-surround and left-surround sources, rather than a single pair, thus creating a "diffuse" sound-field effect as is important for THX -specified systems), and (c) the virtualisation of Dolby Surround (RTM) for headphones, creating a pair of decorrelated rear channels from the single, provided surround channel.

Finally, the accompanying abstract is incorporated herein by reference.

CLAIMS

1. A method of generating a second audio signal from a first audio signal, for use in synthesising a three dimensional sound field, the second audio signal being sufficiently decorrelated from the first audio signal that it can be perceived to be independent of said first audio signal by a listener in use, the method including or consisting of the following steps:- a) deriving from the first signal a first delayed audio signal; b) multiplying this first delayed audio signal by a gain factor between zero and minus one to give a first delayed gain-adjusted audio signal; c) deriving from the first signal a second delayed audio signal, having a different delay time from the first delayed audio signal; d) multiplying this second delayed audio signal by a gain factor between zero and plus one to give a second delayed gain-adjusted audio signal; and e) combining said first and said second delayed gain-adjusted signals with the first audio signal, or a time-delayed version of the first audio signal, to provide a second audio signal decorrelated from the first audio signal, characterised in that the first and second delayed audio signals are delayed by time periods which are caused to change from time to time in a random or pseudo-random or quasi-random manner.
2. A method as claimed in claim 1 in which the said gain factors of the first and second delayed gain-adjusted signals sum to substantially zero.
3. A method as claimed in claim 1 or claim 2 in which the delay times are caused to change at a frequency of greater than 0.5 Hz.
4. Apparatus for generating a second audio signal from a first audio signal, for use in synthesising a three dimensional sound field, the second audio signal being sufficiently decorrelated from the first audio signal that it can be perceived to be independent of said first audio signal by a listener in use, the apparatus including or consisting of an audio signal delay line having said first audio signal as an input, and a plurality of output tap points each having a respective different delay time within a predetermined range of delay times,

the output signals from each tap point of said plurality being multiplied by a selected gain factor, one of the selected gain factors being negative, the plurality of gain-adjusted output signals from the output tap points being combined with the first audio signal, or a time-delayed version of the first audio signal, to provide a second audio signal decorrelated from the first audio signal, characterized in that the said respective different delay times corresponding to each output tap point are assigned selected fixed values which are arranged to change from time to time to other fixed values within said predetermined range of delay times.

5. Apparatus as claimed in claim 4 in which the plurality of selected gain factors sums to substantially zero.

6. Apparatus as claimed in claim 4 or 5 in which the said selected fixed values are arranged to change to other fixed values at a frequency of greater than 0.5 Hz.

7. Apparatus as claimed in claim 4 - 6, in which the selected fixed values of the respective different delay times are changed in a random or pseudo-random or quasi-random manner.

8. Apparatus as claimed in claim 4 - 7, in which the second audio signal is generated from the first audio signal, or a time-delayed version of the first audio signal, and two output tap points or two sets of output tap points which are cross-faded, the amplitude of the output signal from a given tap point being substantially zero when the delay time of said given tap point is changed from one fixed value to another.

The Patent Office 9

INVESTOR IN PEOPLE

Application No: GB 9920811.8
Claims searched: all

Examiner: Martyn Dixon
Date of search: 9 March 2000

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.R): H4R (RSS,RSX)

Int Cl (Ed.7): H04R (3/00,5/00,5/04); H04S (1/00,3/00,5/00,5/02)

Other: Online: EPODOC,WPI,JAPIO

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
A	WO 98/30064 A	(Central Research Laboratories) see fig 6	
A	US 5555306 A	(Trifield Productions)	
A	US 4625326 A	(US Philips)	

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.